

## **Ligation Method**

### **Field of the Invention**

This application relates to a method of ligating two or more molecules, for example, small organic molecules, labels, peptides etc. In particular it relates to a method of ligating a peptide, such as ligation of a synthetic peptide to a recombinant peptide.

### **Background to the Invention**

Protein engineering methodologies have proven to be invaluable for generating protein based tools for application in basic research, diagnostics, drug discovery and as protein therapeutics. The ability to manipulate the primary structure of a protein in a controlled manner opens up many new possibilities in the biological and medical sciences. As a consequence, there is a concerted effort on developing methodologies for the site-specific modification of proteins and their subsequent application.

1

2 The two main approaches to generating proteins are  
3 through recombinant methods or chemical synthesis.

4 To date, the two methods have proved to be  
5 complementary; recombinant methodologies enable  
6 proteins of any size to be generated but in general  
7 they are restricted to the assembly of the  
8 proteinogenic amino acids. Thus, in general, the  
9 introduction of labels and probes into recombinant  
10 proteins has to be implemented post-translationally  
11 and does not allow modifications to the protein  
12 backbone.

13

14 The most common methods for labelling a recombinant  
15 protein use an amino or a thiol reactive version of  
16 the label that will covalently react with a lysine  
17 side chain / N<sup>α</sup> amino group or a cysteine side chain  
18 within the protein respectively. For such labelling  
19 methods to be site-specific, an appropriate  
20 derivative of the protein must be engineered to  
21 contain a unique reactive functionality at the  
22 position to be modified. This requires all the other  
23 naturally occurring reactive functionalities within  
24 the primary sequence to be removed through amino  
25 acid mutagenesis. In the case of protein amino  
26 functionalities, this is essentially impossible due  
27 to the abundance of lysine residues within proteins  
28 and the presence of the amino functionality at the  
29 N-terminus of the sequence. Likewise, for cysteine  
30 this process is laborious and is often detrimental  
31 to the function of the protein.

32

1 The production of proteins having site-specific  
2 modifications and/or labels is more readily  
3 achievable using chemical synthesis methods. The  
4 chemical synthesis of proteins enables multiple  
5 modifications to be incorporated into both side-  
6 chain and backbone moieties of the protein in a  
7 site-specific manner, but, in general, the maximum  
8 size of sequence that can be synthesised and  
9 isolated is circa 50 - 100 amino acids.

10

### 11 **Protein Ligation**

12 A further approach to the generation of proteins is  
13 protein / peptide ligation. In this approach  
14 mutually reactive chemical functionalities  
15 (orthogonal to the chemistry of the naturally  
16 occurring amino acids i.e. which react by mutually  
17 exclusive chemistries compared to the reactions of  
18 the reactive moieties of the naturally occurring  
19 amino acids) are incorporated at the N- and C-  
20 termini of unprotected polypeptide fragments such  
21 that when they are mixed, they react in a  
22 chemoselective manner to join the two sequences  
23 together (Cotton GJ and Muir TW. Chem.Biol., 1999,  
24 6, R247-R254). The principle of chemical ligation is  
25 shown schematically in Figure 1.

26

27 A number of chemistries have been utilised for the  
28 ligation of two synthetic peptides where a diverse  
29 range of different chemical functionalities can be  
30 incorporated into the termini of polypeptides using  
31 solid phase peptide synthesis. These include the  
32 reaction between a thioacid and bromo- alkyl to

1 form a thioester (Schnolzer M and Kent SBH, *Science*,  
2 1992, 256, 221-225), reaction of an aldehyde with an  
3 N-terminal cysteine or threonine to form  
4 thiazolidine or oxazolidine respectively (Liu C-F  
5 and Tam J P. *Proc. Natl. Acad. Sci. USA*, 1994, 91,  
6 6584 - 6588), reaction between a hydrazide and an  
7 aldehyde to form a hydrazone (Gaertner HF et al, et  
8 al *Bioconj. Chem.*, 1992, 3, 262 - 268) reaction of  
9 an aminoxy group and an aldehyde to form an oxime  
10 (Rose K. *J. Am. Chem. Soc.*, 1994, 116, 30-33),  
11 reaction of azides and aryl phosphines to form an  
12 amide bond (Staudinger ligation) (Nilsson BL,  
13 Kiessling LL, and Raines RT. *Org. Lett.*, 2001, 3, 9-  
14 12, Kiick et al *Proc. Natl. Acad. Sci. USA*, 2002,  
15 99, 19-24) , and the reaction of a peptide C-  
16 terminal thioester and an N-terminal cysteine  
17 peptide to form a native amide bond (Dawson et al.  
18 *Science*, 1994, 266, 776) (Native chemical ligation  
19 US6184344, EP 0832 096 B1). This native chemical  
20 ligation method is an extension of studies by  
21 Wieland and coworkers who showed that the reaction  
22 of ValSPH and CysOH in aqueous buffer yielded the  
23 dipeptide ValCysOH (Wieland T et al, . *Liebigs Ann.*  
24 *Chem.*, 1953, 583, 129-149).

25

26 Although the native chemical ligation method has  
27 proved popular, it requires an N-terminal cysteine  
28 containing peptide for the reaction and thus, if a  
29 cysteine is not present at the appropriate position  
30 in the protein, a cysteine needs to be introduced at  
31 the ligation site. However, the introduction of  
32 extra thiol groups into a protein sequence may be

1 detrimental to its structure / function, especially  
2 since cysteine has a propensity to form disulfide  
3 bonds which may disrupt the folding pathway or  
4 compromise the function of the folded protein.  
5

6 As a consequence of the difficulties and problems  
7 associated with known ligation techniques, the  
8 ligation of two synthetic fragments generally only  
9 enables proteins of circa 100 - 150 amino acids to  
10 be chemically synthesised. Although larger proteins  
11 have been synthesised by ligating together more than  
12 two fragments, this has proved to be technically  
13 difficult (Camarero et al. *J. Pept. Res.*, 1998, 54,  
14 303-316, Canne LE et al, *J. Am. Chem. Soc.*, 1999,  
15 121, 8720-8727).  
16

#### 17 **Protein semi-synthesis**

18  
19 protein ligation technologies that enable both  
20 synthetic and recombinantly derived protein  
21 fragments to be joined together have been described.  
22 This enables large proteins to be constructed from  
23 combinations of synthetic and recombinant fragments,  
24 allowing proteins to be site-specifically modified  
25 with both natural and unnatural entities. By  
26 utilising such so-called protein semi-synthesis,  
27 many different synthetic moieties can be site-  
28 specifically incorporated at multiple different  
29 sites within a target protein.  
30

31 In order to utilise recombinant proteins in ligation  
32 strategies the recombinant fragments must contain

1 the appropriate reactive functionalities to  
2 facilitate ligation. One approach to introduce a  
3 unique reactive functionality into a recombinant  
4 protein has been through the periodate oxidation of  
5 N-terminal serine containing sequences. Such  
6 treatment converts the N-terminal serine into a  
7 glyoxyl moiety, which contains an N-terminal  
8 aldehyde. Synthetic hydrazide containing peptides  
9 have then been ligated to the N-terminus of these  
10 proteins in a chemoselective manner through  
11 hydrazone bond formation with the protein N-terminal  
12 glyoxyl group (Gaertner HF et al, et al Bioconj.  
13 Chem., 1992, 3, 262 - 268, Gaertner HF, et al. *J.*  
14 *Biol. Chem.*, 1994, 269, 7224-7230). Another approach  
15 has been to generate recombinant proteins with N-  
16 terminal cysteine residues. Synthetic peptides  
17 containing C-terminal thioesters have then been  
18 site-specifically attached to the N-terminus of  
19 these proteins via amide bond formation in a manner  
20 analogous to 'native chemical ligation' (Cotton GJ  
21 and Muir TW. *Chem. Biol.*, 2000, 7, 253-261). However  
22 as with the ligation of synthetic peptides using  
23 native chemical ligation techniques, the technology  
24 requires a cysteine to be introduced at the ligation  
25 site if the primary sequence does not contain one at  
26 the appropriate position.

27

## 28 **Protein Splicing Techniques**

29

30 Recently technologies have been developed which  
31 enable recombinant proteins containing C-terminal  
32 thioester groups to be generated. The C-terminal

1 thioester functionality provides a unique reactive  
2 chemical group within the protein that can be  
3 utilised for protein ligation. Recombinant C-  
4 terminal thioester proteins are produced by  
5 manipulating a naturally occurring biological  
6 phenomenon known as protein splicing (Paulus H. *Annu*  
7 *Rev Biochem* 2000, 69, 447-496). Protein splicing is  
8 a post-translational process in which a precursor  
9 protein undergoes a series of intramolecular  
10 rearrangements which result in precise removal of an  
11 internal region, referred to as an intein, and  
12 ligation of the two flanking sequences, termed  
13 exteins (Figure 2). While there are generally no  
14 sequence requirements in either of the exteins,  
15 inteins are characterised by several conserved  
16 sequence motifs and well over a hundred members of  
17 this protein domain family have now been identified.

18  
19 The first step in protein splicing involves an N→S  
20 (or N→O) acyl shift in which the N-extein unit is  
21 transferred to the sidechain SH or OH group of a  
22 conserved Cys/Ser/Thr residue, always located at the  
23 immediate N-terminus of the intein. Insights into  
24 this mechanism have led to the design of a number of  
25 mutant inteins which can only promote the first step  
26 of protein splicing (Chong et al *Gene*. 1997, 192,  
27 271-281, (Noren et al., *Angew. Chem. Int. Ed. Engl.*,  
28 2000, 39, 450-466). Proteins expressed as in frame  
29 N-terminal fusions to one of these engineered  
30 inteins can be cleaved by thiols via an  
31 intermolecular transthioesterification reaction, to  
32 generate the recombinant protein C-terminal

1 thioester derivative (Figure 3) (Chong et al *Gene*.  
2 1997, 192, 271-281, (Noren et al., *Angew. Chem. Int.*  
3 *Ed. Engl.*, 2000, 39, 450-466) (New England Biolabs  
4 Impact System WO 00/18881, WO 0047751). Peptide  
5 sequences containing an N-terminal cysteine residue  
6 can then be specifically ligated to the C-termini of  
7 such recombinant C-terminal thioester proteins (Muir  
8 et al *Proc. Natl. Acad. Sci. USA.*, 1998, 95, 6705-  
9 6710, Evans Jr et al. *Prot. Sci.*, 1998, 7, 2256-  
10 2264) , in a procedure termed expressed protein  
11 ligation (EPL) or intein-mediated protein ligation  
12 (IPL).

13

14 The chemoselective ligation of N-terminal cysteine  
15 containing peptides to C-terminal thioester  
16 containing peptides, be they synthetic or  
17 recombinant, is performed typically at slightly  
18 basic pH and in the presence of a thiol cofactor.  
19 The strategy also requires a cysteine to be  
20 introduced at the ligation site, if one is not  
21 suitably positioned within the primary sequence.  
22 These requirements of this ligation approach have  
23 the potential to alter the structure and / or  
24 function of both the protein ligation product and  
25 the initial reactants.

26

27 For example, the chemokine RANTES is unstable in a  
28 buffer of 100 mM NaCl, 100 mM sodium phosphate pH  
29 7.4 containing 100 mM 2-mercaptoethanesulfonic acid  
30 (MESNA); a buffer typically used for the ligation of  
31 C-terminal thioester molecules to N-terminal  
32 cysteine containing molecules (expressed protein



1 ligation and native chemical ligation). RANTES  
2 contains two disulphide bonds critical for  
3 maintaining the structure and function of the  
4 protein. In the typical ligation buffer described  
5 above, the folded protein was found to be converted  
6 within 48 hours to a mixture of the reduced protein  
7 and MESNA protein adducts. The majority of the  
8 protein mixture subsequently formed a precipitate,  
9 presumably reflecting the unfolded nature of these  
10 species (Cotton, unpublished).

11  
12 Accordingly, the inventors believe that ligation  
13 reactions that require thiol containing buffers are,  
14 in general, not suitable for maintaining the  
15 integrity of disulphide bond containing proteins,  
16 such as antibodies, antibody fragments and antibody  
17 domains, cytokines, growth factors etc. Thus there  
18 is a requirement for ligation approaches that are  
19 typically performed in the absence of thiols. For  
20 example, when monitored over a number of days, it  
21 was found that RANTES was stable in 100 mM NaCl, 100  
22 mM sodium phosphate buffer pH 7.4 and 100 mM sodium  
23 acetate buffer pH 4.5 (inventor's unpublished  
24 results). Ligation reactions that can be performed  
25 under such conditions should therefore be applicable  
26 for both disulphide and non-disulphide containing  
27 proteins.

28

### 29 **Protein labelling**

30

31 Historically protein ligation means the joining  
32 together of two peptide / protein fragments but this

1 is synonymous with protein labelling whereby the  
2 label is a peptide or derivatised peptide. Equally  
3 if a small non-peptidic synthetic molecule contains  
4 the necessary reactive chemical functionality for  
5 protein ligation, then ligation of the synthetic  
6 molecule directly to either the N- or C- termini of  
7 the protein affords site-specific labelling of the  
8 protein. Thus technologies developed for the  
9 ligation of protein fragments can also be used for  
10 the direct labelling of either the N- or C- termini  
11 of peptides or proteins in a site - specific manner  
12 irrespective of their sequence.

13

14 Recombinant proteins containing N-terminal glyoxyl  
15 functions (generated through periodate oxidation of  
16 the corresponding N-terminal serine protein) have  
17 been site-specific N-terminally labelled through  
18 reaction with hydrazide or aminoxy derivatives of  
19 the label (Geoghegan KF and Stroh JG. *Bioconj Chem.*,  
20 1992, 3, 138-146, Alouni S et al. *Eur. J. Biochem.*,  
21 1995, 227, 328 - 334). Also recombinant proteins  
22 containing N-terminal cysteine residues have been N-  
23 terminally labelled through reaction with labels  
24 containing thioester functionalities, the label  
25 being the acyl substituent of the thioester (Schuler  
26 B and Pannell LK. *Bioconjug. Chem.*, 2002, 13, 1039-  
27 43) and aldehyde functionalities (Zhao et al.  
28 *Bioconj. Chem.*, 1999, 10, 424-430) to form amides  
29 and thiazolidines respectively.

30

31 Though a number of methods for ligation of proteins  
32 exist each one has its potential drawbacks. There

1 is thus a need for novel ligation methodologies,  
2 especially those that are compatible with both  
3 synthetic and recombinant fragments, and which may  
4 be used in the ligation of disulphide bond  
5 containing proteins as well as non disulphide bond  
6 containing proteins, which will complement the  
7 existing technologies and add another string to the  
8 protein engineer's bow.

9

#### 10 **Summary of the Invention**

11

12 The present inventors have overcome a number of  
13 problems associated with the prior art and have  
14 developed a new method for ligating peptide  
15 molecules which overcomes a number of the problems  
16 of the prior art.

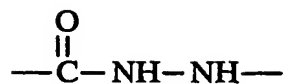
17

18 Accordingly, in a first aspect of the present  
19 invention, there is provided a method of producing  
20 an oligopeptide product, the method comprising the  
21 steps:

- 22 a) providing a first oligopeptide, the first
- 23 oligopeptide having a reactive moiety,
- 24 b) providing a second oligopeptide, the second
- 25 oligopeptide having an activated ester moiety
- 26 c) allowing the reactive moiety of the first
- 27 oligopeptide to react with the activated ester
- 28 moiety of the second oligopeptide to form an
- 29 oligopeptide product, in which the first and second
- 30 oligopeptides are linked via a linking moiety having
- 31 Formula I, Formula II or Formula III.

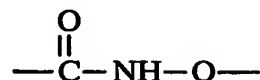
32

1           Formula I



2

3           Formula II



4

5           Formula III



6

7

8

9           In preferred embodiments, in step (c), where said  
10           oligopeptides are linked via a linking moiety having  
11           Formula II and where said activated ester moiety of  
12           step (b) is not a thioester, said activated ester is  
13           a terminal activated ester moiety.

14

15           In further preferred embodiments of the invention,  
16           said linking moieties are linked via a linking  
17           moiety having Formula I or Formula III.

18

19           Unless the context demands otherwise, the terms  
20           peptide, oligopeptide, polypeptide and protein are  
21           used interchangeably.

22

23           The activated ester moiety of the first oligopeptide  
24           may be any suitable activated ester moiety, such as  
25           a thioester moiety, a phenolic ester moiety, an

1 hydroxysuccinimide moiety, or an O-acylisourea  
2 moiety.

3

4 In preferred embodiments of the invention, the  
5 activated ester moiety is a thioester moiety. Any  
6 suitable thioester peptides wherein the peptide is  
7 the acyl substituent of the thioester may be used in  
8 the present invention (Figure 4).

9

10 Such thioester peptides may be synthetically or  
11 recombinantly produced. The skilled person is well  
12 aware of methods known in the art for generating  
13 synthetic peptide thioesters. For example, synthetic  
14 peptide thioesters may be produced via synthesis on  
15 a resin that generates a C-terminal thioester upon  
16 HF cleavage (Hojo et al, Bull. Chem. Soc. Jpn.,  
17 1993, 66, 2700-2706). Further, the use of 'safety  
18 catch' linkers has proved to be popular for  
19 generating C-terminal thioesters through thiol  
20 induced resin cleavage of the assembled peptide  
21 (Shin Y et al, J. Am. Chem. Soc., 1999, 121, 11684-  
22 11689).

23

24 Moreover, recently technologies have been developed  
25 which enable recombinant C-terminal thioester  
26 proteins to be generated. Recombinant C-terminal  
27 thioester proteins may be produced by manipulating a  
28 naturally occurring biological phenomenon known as  
29 protein splicing. As described above, protein  
30 splicing is a post-translational process in which a  
31 precursor protein undergoes a series of  
32 intramolecular rearrangements which result in

1 precise removal of an internal region, referred to  
2 as an intein, and ligation of the two flanking  
3 sequences, termed exteins.

4  
5 As described above, a number of mutant inteins which  
6 can only promote the first step of protein splicing  
7 have been designed (Chong et al *Gene*. 1997, 192,  
8 271-281, Noren et al., *Angew. Chem. Int. Ed. Engl.*,  
9 2000, 39, 450-466). Proteins expressed as in frame  
10 N-terminal fusions to one of these engineered  
11 inteins can be cleaved by thiols via an  
12 intermolecular transthioesterification reaction, to  
13 generate the recombinant protein C-terminal  
14 thioester derivative (Chong et al *Gene*. 1997, 192,  
15 271-281, Noren et al., *Angew. Chem. Int. Ed. Engl.*,  
16 2000, 39, 450-466) (New England Biolabs Impact  
17 System WO 00/18881, WO 0047751). Such protein  
18 thioesters may be used in the methods of the  
19 invention (See Figure 3).

20  
21 Accordingly, in a preferred aspect of the present  
22 invention, in step (b), the second oligopeptide is  
23 generated by thiol reagent induced cleavage of an  
24 intein fusion protein.

25  
26 Accordingly, in a second aspect of the present  
27 invention, there is provided a method of producing  
28 an oligopeptide product, the method comprising the  
29 steps:

30 a) providing a first oligopeptide, the first  
31 oligopeptide having a reactive moiety,

1     b)     (i) providing a precursor oligopeptide  
2     molecule, the precursor oligopeptide molecule  
3     comprising a precursor second oligopeptide fused N-  
4     terminally to an intein domain  
5     (ii) allowing thiol reagent dependent cleavage of  
6     the precursor molecule to generate a second  
7     oligopeptide molecule, said second oligopeptide  
8     molecule having a thioester moiety at its C-terminus  
9     c) allowing the reactive moiety of the first  
10    oligopeptide to react with the second oligopeptide  
11    molecule to form an oligopeptide product, in which  
12    the first and second oligopeptides are linked via a  
13    linking moiety having Formula I, II or III.

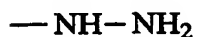
14

15    The reactive moiety of the first oligopeptide may be  
16    any suitable reactive moiety. In preferred  
17    embodiments of the invention, the reactive moiety is  
18    a hydrazine moiety, an amino-oxy moiety or a  
19    hydrazide moiety having general formula IV, V or VI  
20    respectively.

21

22    Formula IV

23



24

25

26    Formula V



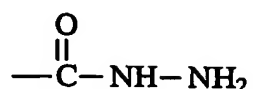
27

28

29

30    Formula VI

31



1  
2 For example, in a particular preferred embodiment,  
3 the reactive moiety has Formula IV and, in the  
4 oligopeptide product produced by the method of the  
5 invention, the first and second oligopeptides are  
6 linked via a linking moiety having Formula I.

7  
8 In a further preferred embodiment, the reactive  
9 moiety has Formula V and, in the oligopeptide  
10 product produced by the method of the invention, the  
11 first and second oligopeptides are linked via a  
12 linking moiety having Formula II.

13  
14 In another preferred embodiment, the reactive moiety  
15 has Formula VI and, in the oligopeptide product  
16 produced by the method of the invention, the first  
17 and second oligopeptides are linked via a linking  
18 moiety having Formula III.

19  
20 As described above, the first oligopeptide comprises  
21 a reactive moiety, which, in preferred embodiments,  
22 may be a hydrazine moiety (e.g. Formula IV), an  
23 amino-oxy moiety (e.g. Formula V) or an hydrazide  
24 moiety (e.g. Formula VI).

25  
26 A particular advantage of the ligation method of the  
27 invention is that it may be performed in the absence  
28 of thiols. This enables efficient ligation of  
29 proteins/peptides comprising disulphide bonds as  
30 well as of proteins without such bonds.

31



1 Accordingly, in an embodiment of the first and  
2 second aspects of the invention, at least one of the  
3 first and second oligopeptides comprises one or more  
4 disulphide bonds.

5  
6 Hydrazine, hydrazide or aminooxy containing  
7 derivatives of synthetic oligopeptides may be  
8 readily produced using known methods, for example,  
9 solid phase synthesis techniques.

10  
11 Further, the present inventors have also found that  
12 proteins fused N-terminal to an intein domain can be  
13 cleaved from the intein by hydrazine treatment in a  
14 selective manner to liberate the desired protein as  
15 its corresponding hydrazide derivative (for example,  
16 see Figure 5).

17  
18 Accordingly, in further preferred embodiments of the  
19 invention, the first oligopeptide is generated by  
20 reaction of hydrazine with an oligopeptide molecule  
21 comprising the first oligopeptide fused N-terminal  
22 to an intein domain.

23  
24 Indeed the discovery that such protein hydrazides  
25 may be produced by means of such a reaction forms an  
26 independent aspect of the present invention.

27  
28 Accordingly, a third aspect of the invention  
29 provides a method of generating a protein hydrazide,  
30 said method comprising the steps:

31 (a) providing an protein molecule comprising an  
32 oligopeptide fused N-terminal to an intein domain,

1 (b) reacting said protein molecule with hydrazine,  
2 such that the intein domain is cleaved from the  
3 oligopeptide to generate a protein hydrazide.

4  
5 Moreover, as well as using such a reaction to  
6 generate a first oligopeptide having a hydrazide  
7 moiety at its C-terminal, the first oligopeptide  
8 thus being available for reaction with the second  
9 oligopeptide having the activated ester moiety, the  
10 present invention further extends to a "one-step"  
11 process for ligating two peptides to generate an  
12 oligopeptide product.

13  
14 This may be achieved by reacting a suitable protein  
15 linked N-terminal to an intein directly with a  
16 polypeptide having a hydrazine, hydrazide or amino-  
17 oxy moiety.

18  
19 Accordingly, in a fourth aspect, the invention  
20 provides a method of producing an oligopeptide  
21 product, the method comprising the steps:

22 a) providing a first oligopeptide, the first  
23 oligopeptide having a reactive moiety, wherein the  
24 reactive moiety is a hydrazine moiety, a hydrazide  
25 moiety or an amino-oxy moiety;  
26 (i) providing a precursor oligopeptide molecule, the  
27 precursor oligopeptide molecule comprising a second  
28 oligopeptide fused N-terminally to an intein domain;  
29 (c) allowing the reactive moiety of the first  
30 oligopeptide to react with the precursor  
31 oligopeptide molecule to form an oligopeptide  
32 product, in which the first and second oligopeptides

1 are linked via a linking moiety having Formula I,  
2 Formula II or Formula III.

3  
4 The ligation technology of the present invention can  
5 thus utilise both synthetic and recombinant proteins  
6 and peptides. It thus enables the ligation of two or  
7 more synthetic peptides, the ligation of two or more  
8 recombinant peptides or the ligation of at least one  
9 synthetic peptide with at least one recombinant  
10 peptide.

11  
12 Moreover, as well as providing a novel method of  
13 ligating peptides, the present invention may be used  
14 for the labelling of synthetic or recombinant  
15 peptides.

16  
17 Accordingly, in a fifth aspect of the present  
18 invention, there is provided a method of labelling  
19 an oligopeptide, the method comprising the steps:  
20 a) providing a label molecule, the label molecule  
21 having a reactive moiety,  
22 b) providing the oligopeptide, the oligopeptide  
23 having an activated ester moiety  
24 c) allowing the reactive moiety of the label  
25 molecule to react with the activated ester moiety of  
26 the oligopeptide to form the labelled oligopeptide,  
27 in which the label molecule and the oligopeptide are  
28 linked via a linking moiety having Formula I,  
29 Formula II or Formula III as defined above,

30  
31 In preferred embodiments, in step (c), where said  
32 label molecule and the oligopeptide are linked via a

1 linking moiety having Formula II and where said  
2 activated ester moiety of step (b) is not a  
3 thioester, said activated ester is a terminal  
4 activated ester moiety.

5

6 In a preferred aspect of the present invention, in  
7 step (b) the oligopeptide is generated by thiol  
8 induced cleavage of an intein fusion protein.

9

10 Accordingly, in a sixth aspect of the present  
11 invention, there is provided a method of labelling  
12 an oligopeptide, the method comprising the steps:

13 a) providing a label molecule, the label molecule  
14 having a reactive moiety,

15 c) (i) providing a precursor oligopeptide  
16 molecule, the precursor oligopeptide molecule  
17 comprising a precursor oligopeptide fused N-  
18 terminally to an intein domain

19 (ii) allowing thiol reagent dependent cleavage of  
20 the precursor molecule to generate an oligopeptide  
21 molecule, said oligopeptide molecule having a  
22 thioester moiety at its C-terminus

23 c) allowing the reactive moiety of the label  
24 molecule to react with the oligopeptide to form the  
25 labelled oligopeptide, in which the label molecule  
26 and the oligopeptide are linked via a linking moiety  
27 having Formula I, II or III.

28

29 Alternatively, a label molecule having a terminal  
30 activated ester moiety may be used to label an  
31 oligopeptide having a reactive moiety. Thus, in a  
32 seventh aspect of the invention, there is provided a

1 method of labelling an oligopeptide, the method  
2 comprising the steps:  
3 a) providing a label molecule, the label molecule  
4 having an activated ester moiety of which the label  
5 is the acyl substituent,  
6 b) providing the oligopeptide, the oligopeptide  
7 having a reactive moiety  
8 c) allowing the activated ester moiety of the label  
9 molecule to react with the reactive moiety of the  
10 oligopeptide to form the labelled oligopeptide, in  
11 which the label molecule and the oligopeptide are  
12 linked via a linking moiety having Formula I,  
13 Formula II or Formula III

14 wherein, in step (c), where said label molecule  
15 and the oligopeptide are linked via a linking moiety  
16 having Formula II and where said activated ester  
17 moiety of step (b) is not a thioester, said  
18 activated ester is a terminal activated ester  
19 moiety.  
20

21 As with the ligation technology, an oligopeptide  
22 present as a precursor molecule linked to an intein  
23 molecule may be labelled directly. Thus, an eighth  
24 aspect of the present invention provides a method of  
25 labelling an oligopeptide, the method comprising the  
26 steps:

27 a) providing a label molecule, the label molecule  
28 having a reactive moiety,  
29 b) providing a precursor oligopeptide molecule,  
30 the precursor oligopeptide molecule comprising an  
31 oligopeptide fused N-terminally to an intein domain,

1 c) allowing the reactive moiety of the label  
2 molecule to react with the precursor oligopeptide  
3 molecule to form a labelled oligopeptide product, in  
4 which the label molecule and the oligopeptide are  
5 linked via a linking moiety having Formula I,  
6 Formula II or Formula III as defined above.

7  
8 Any suitable label molecule known to the skilled  
9 person may be used in methods of the invention. The  
10 choice of label will depend on the use to which the  
11 labelled peptide is to be put. For example labels  
12 which may be used in the methods of the invention  
13 may include fluorophores, crosslinking reagents,  
14 spin labels, affinity probes, imaging reagents, for  
15 example radioisotopes, chelating agents such as  
16 DOTA, polymers such as PEG, lipids, sugars, cytotoxic  
17 agents, and solid surfaces and beads.

18  
19 In particular embodiments of the fifth, sixth, and  
20 seventh aspects of the invention, at least one of  
21 the label and oligopeptides comprises one or more  
22 disulphide bonds.

23  
24 The methods of the invention are particularly useful  
25 in the ligation of peptides, in particular the  
26 ligation of peptides, which, using conventional  
27 ligation techniques, would require various  
28 protecting groups. The inventors have shown that  
29 the methods of the invention may be performed under  
30 pH conditions in which only the reactive moieties  
31 will react.

32

1 In preferred embodiments of the first and second and  
2 in preferred embodiments of the fourth to eighth  
3 aspects of the invention, step (c) of the method is  
4 performed at a pH in the range pH 4.0 to pH 8.5,  
5 preferably pH 4.0 to 8.0, for example, pH 4.0 to  
6 7.5, more preferably in the range pH 5.0 to pH 8.0,  
7 more preferably in the range pH 6.0 to pH 7.5, most  
8 preferably in the range pH 6.5 to pH 7.5.

9  
10 For example, the inventors have demonstrated that  
11 synthetic peptide C-terminal thioesters specifically  
12 react with hydrazine under aqueous conditions at pH  
13 6.0 to form the corresponding peptide hydrazide.  
14 This allows ligation methods as described herein to  
15 be performed at pH 6.0, without the need for a  
16 potentially harmful thiol cofactor (useful if either  
17 fragment or final construct is thiol sensitive) and  
18 does not lead to the introduction of potentially  
19 reactive side-chain groups (such as a thiol) into  
20 the protein. Similarly, the inventors have  
21 demonstrated that synthetic peptide C-terminal  
22 thioesters specifically react with hydroxylamine  
23 under aqueous conditions at pH 6.0 and pH 6.8 to  
24 form the corresponding peptide hydroxamic acid.  
25 In addition, as described below, the inventors have  
26 demonstrated that both synthetic peptide C-terminal  
27 thioesters and recombinant protein C-terminal  
28 thioesters specifically react with O-  
29 methylhydroxylamine under aqueous conditions at pH  
30 7.5, to form the corresponding C-terminal N-methoxy  
31 amide derivatives. This allows ligation methods as

1 described herein to be performed at pH 7.5, without  
2 the need for a potentially harmful thiol cofactor.

3  
4 Peptides and proteins that contain thioester  
5 groups (where the peptide is the acyl substituent of  
6 the thioester) can be reacted with hydrazine,  
7 hydrazide or aminooxy derivatives of a label or a  
8 peptide to afford site-specific labelling and  
9 chemoselective ligation respectively (see, for  
10 example, figures 4 and 5).

11  
12 In an analogous fashion, peptides that contain  
13 hydrazine, hydrazide or aminooxy groups can be  
14 reacted with thioester derivatives of a label or a  
15 peptide to afford site-specific labelling and  
16 chemoselective ligation respectively (see, for  
17 example, figures 4 and 5).

18  
19 Furthermore, having demonstrated that recombinant  
20 protein hydrazides can be generated by cleavage of  
21 protein-intein fusions with hydrazine, the inventors  
22 have shown that such protein hydrazides may be  
23 ligated by reaction of the hydrazide moiety with  
24 reactive groups other than activated ester moieties,  
25 for example an aldehyde functionality or a ketone  
26 functionality. For example, as described below, the  
27 inventors have shown that a pyruvoyl derivative of a  
28 synthetic peptide can be chemoselectively ligated to  
29 the C-terminus of recombinant protein hydrazides  
30 using the described approach, and in an analogous  
31 fashion, a pyruvoyl derivative of fluorescein was  
32 used to site-specifically label the C-terminus of



1 recombinant protein hydrazides using the described  
2 approach.

3  
4 This aspect of the invention provides a further  
5 novel method of ligating a recombinant peptide to a  
6 second peptide or indeed a label.

7  
8 Thus, a ninth aspect of the invention provides a  
9 method of producing an oligopeptide product, the  
10 method comprising the steps:

- 11 a) providing a first oligopeptide, the first
- 12 oligopeptide having an aldehyde or ketone moiety,
- 13 b) providing a precursor oligopeptide molecule,
- 14 the precursor oligopeptide molecule comprising a
- 15 second oligopeptide fused N-terminally to an intein
- 16 domain,
- 17 c) reacting said precursor oligopeptide molecule
- 18 with hydrazine to generate an oligopeptide molecule
- 19 comprising an intermediate oligopeptide, said
- 20 intermediate oligopeptide having a C-terminal
- 21 hydrazide moiety,
- 22 d) allowing the aldehyde or ketone moiety of the
- 23 first oligopeptide to react with the hydrazide
- 24 moiety of the intermediate oligopeptide molecule to
- 25 form an oligopeptide product, in which first
- 26 oligopeptide and the second oligopeptide are linked
- 27 via a hydrazone linking moiety.

28  
29 An example of this aspect is shown in Figure 6.

30

1 A tenth aspect of the invention provides a method of  
2 labelling an oligopeptide, the method comprising the  
3 steps:

4 a) providing a label molecule, the label molecule  
5 having a aldehyde or ketone moiety,

6 b) providing a precursor oligopeptide molecule,  
7 the precursor oligopeptide molecule comprising a  
8 first oligopeptide fused N-terminally to an intein  
9 domain,

10 c) reacting said precursor oligopeptide molecule  
11 with hydrazine to generate an oligopeptide molecule  
12 comprising an intermediate oligopeptide, said  
13 intermediate oligopeptide having a terminal  
14 hydrazide moiety,

15 d) allowing the aldehyde or ketone moiety of the  
16 label molecule to react with the hydrazide moiety of  
17 the intermediate oligopeptide molecule to form a  
18 labelled oligopeptide product, in which the label  
19 molecule and oligopeptide are linked via a hydrazone  
20 linking moiety.

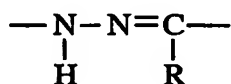
21

22 In preferred embodiments of the ninth and tenth  
23 aspects of the invention, the hydrazone moiety has  
24 Formula VII:

25

26

27



28

29 where R is H or any substituted or unsubstituted,  
30 preferably unsubstituted, alkyl group.

31

1 In preferred aspects of the ninth and tenth aspects  
2 of the invention, the method is performed at a pH in  
3 the range pH 1.0 to pH 7.0, preferably pH 1.0 to pH  
4 6.0, more preferably in the range pH 2.0 to pH 5.5,  
5 most preferably in the range pH 2.0 to pH 4.5.

6  
7 In a particular embodiment of the ninth and tenth  
8 aspects of the invention, the aldehyde or ketone  
9 containing moiety of the oligopeptide or of the  
10 label is an  $\alpha$ -diketone group or an  $\alpha$ -keto aldehyde  
11 group.

12  
13 In a eleventh aspect of the present invention, there  
14 is provided an oligopeptide product produced using a  
15 method of the invention.

16  
17 In an twelfth aspect, there is provided a labelled  
18 oligopeptide comprising an oligopeptide labelled  
19 according to a method of the invention.

20  
21 Preferred features of each aspect of the invention  
22 are as for each of the other aspects mutatis  
23 mutandis.

24  
25 The invention will now be described further in the  
26 following non-limiting examples with reference made  
27 to the accompanying drawings in which:

28  
29 Figure 1 illustrates schematically the general  
30 principle of chemical ligation.

31

1     Figure 2 illustrates schematically the mechanism of  
2     protein splicing.

3

4     Figure 3 illustrates the generation of recombinant  
5     C-terminal thioester proteins.

6

7     Figure 4 illustrates ligation of protein and peptide  
8     thioesters with hydrazine and aminooxy containing  
9     entities, such as labels, peptides and proteins.

10

11     Figure 5 illustrates the generation of synthetic and  
12     recombinant peptide hydrazides for ligation with  
13     thioester containing molecules. Note the peptide or  
14     label is is the acyl substituent of the thioester.

15

16     Figure 6 illustrates the generation of recombinant  
17     peptide hydrazides for ligation with aldehyde and  
18     ketone containing molecules.

19

20     Figure 7 illustrates SDS-PAGE analysis of Grb2-SH2 -  
21     GyrA - CBD (immobilised on chitin beads) treated  
22     with DTT and MESNA. Molecular weight markers (lane  
23     1); purified Grb2-SH2 - GyrA - CBD immobilised on  
24     chitin beads (lane 4). Grb2-SH2 - GyrA - CBD treated  
25     with 100 mM DTT (lanes 5 and 7) or 120 mM MESNA  
26     (lanes 8 and 10). Both the whole reaction slurries  
27     (lanes 5 and 8) and the reaction supernatants (lanes  
28     7 and 10) were analysed.

29

30     Figure 8 illustrates SDS-PAGE analysis of Grb2-SH2 -  
31     GyrA - CBD (immobilised on chitin beads) treated  
32     with hydrazine. Molecular weight markers (lane 1);

1 Purified Grb2-SH2 - GyrA - CBD immobilised on chitin  
2 beads after 20h treatment with phosphate buffer only  
3 (lane 2). Grb2-SH2 - GyrA - CBD treated with 200 mM  
4 hydrazine in phosphate buffer for 20 h. The whole  
5 reaction slurries were analysed.

6  
7 Figure 9 illustrates an ESMS spectrum of the C-  
8 terminal hydrazide derivative of Grb2-SH2.

9  
10 Figure 10 shows SDS-PAGE analysis of the reaction  
11 between synthetic ketone containing peptide CH<sub>3</sub>COCO-  
12 myc with Grb2-SH2 - C-terminal hydrazide and  
13 Cytochrome C. Molecular weight markers (lane 1);  
14 Grb2-SH2 - C-terminal DTT thioester (lane 2).  
15 Reaction between Grb2-SH2 - C-terminal hydrazide and  
16 CH<sub>3</sub>COCO-myc at time points t=0 h (lane 3), t=24 h  
17 (lane 4), t= 48h (lane 5) and t= 72 h (lanes 6).  
18 Reaction between Cytochrome C and CH<sub>3</sub>COCO-myc at  
19 time points t=0 h (lane 7), t=24 h (lane 8), t= 48h  
20 (lane 9) and t= 72 h (lanes 10)

21  
22 Figure 11 shows the structure of CH<sub>3</sub>COCO-Lys(Fl).  
23 The 5-carboxy fluorescein positional isomer is  
24 shown.

25  
26 Figure 12 illustrates SDS-PAGE analysis of the  
27 reaction between CH<sub>3</sub>COCO-Lys(Fl) with Grb2-SH2 C-  
28 terminal hydrazide in 50 mM sodium acetate buffer pH  
29 4.5. Molecular weight markers (lane 1); Grb2-SH2 C-  
30 terminal hydrazide (lane 2). Reaction between Grb2-  
31 SH2 C-terminal hydrazide and CH<sub>3</sub>COCO-Lys(Fl) at

1 time points t=4 h (lane 3), t=24 h (lane 4), t= 48h  
2 (lane 5)

3

4 Figure 13 illustrates SDS-PAGE analysis of the  
5 reaction between CH<sub>3</sub>COCO-Lys(F1) with Cytochrome C  
6 in 100 mM sodium acetate buffer pH 4.5. Molecular  
7 weight markers (lane 1); Cytochrome C (lane 2).

8 Reaction between Cytochrome C and CH<sub>3</sub>COCO-Lys(F1)  
9 at time points t=4 h (lane 3), t=24 h (lane 4), t=  
10 48h (lane 5).

11

12 Figure 14 illustrates SDS-PAGE analysis of the  
13 reaction of CH<sub>3</sub>COCO-Lys(F1) with Grb2-SH2 C-  
14 terminal hydrazide and with Cytochrome C in 50 mM  
15 sodium acetate buffer pH 4.5. (A) total protein stain  
16 of gel. Prior to this coomassie staining (A), the  
17 gel was imaged for green fluorescence (B). Molecular  
18 weight markers (lane 1); Grb2 SH2 C-terminal  
19 hydrazide (lane 2); Reaction between Grb2 SH2 C-  
20 terminal hydrazide and CH<sub>3</sub>COCO-Lys(F1) at time  
21 points t=4 h (lane 3), t=24 h (lane 4), t= 48h (lane  
22 5). Cytochrome C (lane 6); Reaction between  
23 Cytochrome C and CH<sub>3</sub>COCO-Lys(F1) at time points t=4  
24 h (lane 7), t= 24 h (lane 8) and t= 48 h (lanes 9).

25

26 Figure 15 shows SDS-PAGE analysis of the reaction  
27 between CH<sub>3</sub>COCO-Lys(F1) and Grb2 SH2 C-terminal  
28 hydrazide in 40% aqueous acetonitrile containing  
29 0.1% TFA; reaction after 4 h (lane 1), 24 h (lane  
30 2), 48h (lane 3), Grb2 SH2 C-terminal hydrazide  
31 (lane 4).

32

1

2 **Examples**

3

4 **Example 1 -Protein ligation / site specific protein**  
5 **labelling using the reaction of peptide / protein**  
6 **thioesters with compounds containing hydrazine /**  
7 **hydrazide or aminoxy functionalities.**

8

9 *A) Reaction of a peptide C-terminal thioester with*  
10 *100mM hydrazine at pH 6.0*  
11 200 mM sodium phosphate buffer pH 6.0 containing  
12 100mM hydrazine monohydrate (200  $\mu$ L) was added to a  
13 model synthetic C-terminal thioester peptide termed  
14 AS626p1A (200  $\mu$ g) to yield a final peptide  
15 concentration of 317  $\mu$ M. AS626p1A has sequence ARTKQ  
16 TARK(Me)<sub>3</sub> STGGKAPRKQ LATKAARK-COS-(CH<sub>2</sub>)<sub>2</sub>-COOC<sub>2</sub>H<sub>5</sub> (SEQ  
17 ID NO: 1) wherein a single Alanine residue (which  
18 may be any one of the Alanine residues of SEQ ID NO:  
19 1) is substituted by an Arginine residue. The  
20 reaction was incubated at room temperature and  
21 monitored with time by analytical reversed phase  
22 HPLC. Vydac C18 column (5  $\mu$ M, 0.46 x 25 cm). Linear  
23 gradients of acetonitrile water / 0.1% TFA were used  
24 to elute the peptides at a flow rate of 1 mL min<sup>-1</sup>.  
25 Individual peptides eluting from the column were  
26 characterised by electrospray mass spectrometry.

27

28 *B)Reaction of a peptide C-terminal thioester with*  
29 *100mM hydroxylamine at pH 6.0*  
30 200 mM sodium phosphate buffer pH 6.0 containing  
31 100mM hydroxylamine hydrogen chloride (200  $\mu$ L) was

1 added to AS626p1A (200  $\mu$ g) to yield a final peptide  
2 concentration of 317  $\mu$ M. The reaction was incubated  
3 at room temperature and monitored with time by  
4 analytical reversed phase HPLC. Vydac C18 column (5  
5  $\mu$ M, 0.46 x 25 cm). Linear gradients of acetonitrile  
6 water / 0.1% TFA were used to elute the peptides at  
7 a flow rate of 1 mL min<sup>-1</sup>. Individual peptides  
8 eluting from the column were characterised by  
9 electrospray mass spectrometry.

10

11 *C) Reaction of a peptide C-terminal thioester with*  
12 *100 mM hydroxylamine at pH 6.8*  
13 200 mM sodium phosphate buffer pH 6.8 containing  
14 100mM hydroxylamine hydrogen chloride (200  $\mu$ L) was  
15 added to AS626p1A (200  $\mu$ g) to yield a final peptide  
16 concentration of 317  $\mu$ M. The reaction was incubated  
17 at room temperature and monitored with time by  
18 analytical reversed phase HPLC. Vydac C18 column (5  
19  $\mu$ M, 0.46 x 25 cm). Linear gradients of acetonitrile  
20 water / 0.1% TFA were used to elute the peptides at  
21 a flow rate of 1 mL min<sup>-1</sup>. Individual peptides  
22 eluting from the column were characterised by  
23 electrospray mass spectrometry.

24

25 *D) Reaction of a peptide C-terminal thioester with*  
26 *10mM hydroxylamine at pH 6.8*  
27 The procedure as described in C) was repeated,  
28 replacing 100mM hydroxylamine with 10mM  
29 hydroxylamine.

30



1     *E) Reaction of a peptide C-terminal thioester with*  
2     *10mM hydroxylamine at pH 7.5*

3     The procedure as described in D) was repeated, at  
4     pH7.5.

5

6     *F) Reaction of a peptide C-terminal thioester with*  
7     *2mM hydroxylamine at pH 7.5*

8     The procedure as described in E) was repeated,  
9     replacing 10mM hydroxylamine with 2mM hydroxylamine.

10

11    *G) Reaction of a peptide C-terminal thioester with*  
12    *100 mM O-Methylhydroxylamine (NH<sub>2</sub>-O-CH<sub>3</sub>) at pH 7.5*  
13    200 mM sodium phosphate buffer pH 7.5 containing  
14    100mM O-methylhydroxylamine (200 µL) was added to  
15    synthetic C-terminal thioester peptide AS626p1A (200  
16    µg) to yield a final peptide concentration of 317  
17    µM. The reaction was incubated at room temperature  
18    and monitored with time by analytical reversed phase  
19    HPLC. Vydac C18 column (5 µM, 0.46 x 25 cm). Linear  
20    gradients of acetonitrile water / 0.1% TFA were used  
21    to elute the peptides at a flow rate of 1 mL min<sup>-1</sup>.  
22    Individual peptides eluting from the column were  
23    characterised by electrospray mass spectrometry.

24

25    *H) Reaction of a peptide C-terminal thioester with*  
26    *10 mM O-Methylhydroxylamine at pH 7.5*

27    The procedure as described in G) was repeated,  
28    replacing 100 mM O-methylhydroxylamine with 10 mM O-  
29    methylhydroxylamine.

30

1     *I) Reaction of a recombinant protein C-terminal*  
2     *thioester with 100 mM O-Methylhydroxylamine at pH*  
3     *7.5*

4  
5             The C-terminal mercaptoethanesulfonic acid  
6     thioester derivative of recombinant Grb2-SH2, was  
7     generated through cleavage of the fusion protein  
8     Grb2-SH2 - GyrA intein - CBD as described in Example  
9     2 below. This recombinant C-terminal thioester  
10    protein (100 µg) was reacted with 100mM O-  
11    methylhydroxylamine in 200 mM sodium phosphate  
12    buffer pH 7.5 (200 µL). The reaction was incubated  
13    at room temperature and monitored with time by  
14    analytical reversed phase HPLC. Vydac C5 column (5  
15    µm, 0.46 x 25 cm). Linear gradients of acetonitrile  
16    water / 0.1% TFA were used to elute the peptides at  
17    a flow rate of 1 mL min<sup>-1</sup>. Individual peptides  
18    eluting from the column were characterised by  
19    electrospray mass spectrometry.

20

21

## 22     **Results**

23     These examples demonstrate the novel strategy for  
24     protein ligation / site specific protein labelling  
25     of both synthetic and recombinant protein sequences  
26     of the invention using the reaction of peptide /  
27     protein C-terminal thioesters with compounds  
28     containing hydrazine / hydrazide or aminoxy  
29     functionalities.

30

1 As described above, a purified synthetic 27 amino  
2 acid C-terminal thioester peptide (the ethyl 3-  
3 mercaptopropionate thioester derivative) was treated  
4 with hydrazine and hydroxylamine under various  
5 conditions (Table 1).

6  
7 Treatment with 100 mM hydrazine at pH 6.0 formed a  
8 peptide species that eluted earlier than the  
9 starting thioester peptide as analysed by HPLC. This  
10 material was identified as the expected peptide  
11 hydrazide by ESMS: observed mass = 3054 Da, expected  
12 (av. isotope comp) 3053 Da. The reaction of the  
13 peptide C-terminal thioester with hydrazine to form  
14 the peptide hydrazide was monitored with time by  
15 reverse phase HPLC. Only the desired material was  
16 formed with no side product formation even after 3  
17 days. The stability of the peptide hydrazide, under  
18 the reaction conditions, indicates that the reaction  
19 occurs at the C-terminal thioester moiety and is  
20 chemoselective in nature. It also highlights the  
21 applicability of this reaction for protein ligation  
22 and labelling (2 h 70% conversion , 4h >95%  
23 conversion).

24  
25 To ascertain whether aminooxy containing compounds  
26 chemoselectively react with peptide / protein C-  
27 terminal thioesters, to afford protein ligation and  
28 site-specific labelling, a synthetic C-terminal  
29 thioester peptide was treated with hydroxylamine  
30 under various conditions (Table 1).

31

1 A purified synthetic 27 amino acid C-terminal  
2 thioester peptide (ethyl 3-mercaptopropionate  
3 thioester, observed mass 3155 Da) was incubated at  
4 room temperature with different hydroxylamine  
5 concentrations in aqueous buffers of varying pH. In  
6 all cases the peptide C-terminal thioester reacted  
7 to form a single product that eluted earlier than  
8 the starting thioester peptide as analysed by  
9 reverse phase HPLC. This material corresponds to the  
10 expected hydroxamic acid peptide as determined by  
11 ESMS: observed mass = 3052 Da, expected (av. isotope  
12 comp) 3054 Da. The kinetics of the reaction were  
13 monitored using reverse phase HPLC. The peptide C-  
14 terminal thioester was converted to the  
15 corresponding peptide hydroxamic acid in a clean  
16 fashion with no side-product formation. Increasing  
17 the pH of the reaction buffer accelerated the rate  
18 of reaction. For instance, with a concentration of  
19 100mM  $\text{NH}_2\text{OH}$ , on moving from pH 6.0 to pH 6.8 the  
20 percentage product formation after 1h increased from  
21 25% to 91%. The rate of reaction with 100 mM  $\text{NH}_2\text{OH}$   
22 at pH 6.0, was comparable with 10 mM  $\text{NH}_2\text{OH}$  at pH 6.8.

23  
24 The rate of reaction of the peptide C-terminal  
25 thioester with hydroxylamine, to form the  
26 corresponding hydroxamic acid, increases with  
27 increasing pH and decreases with decreasing  $\text{NH}_2\text{OH}$   
28 concentrations. To identify conditions of pH and  
29 reactant concentration suitable for peptide /  
30 protein labelling and ligation, the labelling was  
31 performed under increasing pH and decreasing  $\text{NH}_2\text{OH}$   
32 concentrations.

1  
2 The reaction with 10 mM  $\text{NH}_2\text{OH}$  was 83% complete after  
3 4h at pH 6.8, while at pH 7.5 it was 83% complete  
4 after 2h. On further decreasing the  $\text{NH}_2\text{OH}$   
5 concentration to 2 mM the reaction rate at pH 7.5  
6 decreased markedly, 70% of the starting peptide  $\alpha$ -  
7 thioester being converted to the corresponding  
8 hydroxamic acid after 8hrs. It was noted that a  
9 small amount of a side-product, corresponding in  
10 mass to the peptide acid, was formed during the  
11 reaction. Presumably this was formed by a competing  
12 hydrolysis side reaction at pH 7.5, which was not  
13 observed with 10 mM  $\text{NH}_2\text{OH}$  at pH 7.5 due to the  
14 faster reaction at this higher reactant  
15 concentration.

Reactant	Concentration	pH	Percentage product formation with time				
			1hr	2hr	4hr	8hr	72hr
$\text{NH}_2\text{NH}_2$	100 mM	6.0	-	70	100		
$\text{NH}_2\text{OH}$	100 mM	6.0	25	48.1	76.3	-	100
$\text{NH}_2\text{OH}$	100 mM	6.8	91	100			
$\text{NH}_2\text{OH}$	10 mM	6.8	26	-	83	100	
$\text{NH}_2\text{OH}$	10 mM	7.5	-	82.7	100	100	
$\text{NH}_2\text{OH}$	2 mM	7.5	11.2	17	38	70	80*

18 **Table 1**

19 \*All starting material has reacted with 80%  
20 conversion to the desired product and ~20% to the  
21 hydrolysis side-product.

22  
23 To further investigate the chemoselective reaction  
24 of aminoxy containing compounds with peptide /

1 protein C-terminal thioesters, to afford protein  
2 ligation and site-specific labelling, the synthetic  
3 C-terminal thioester peptide AS626p1 was treated  
4 with *O*-methylhydroxylamine.

5 The purified synthetic 27 amino acid C-terminal  
6 thioester peptide (ethyl 3-mercaptopropionate  
7 thioester, observed mass 3155 Da) was incubated at  
8 room temperature with 100mM *O*-methylhydroxylamine in  
9 200 mM sodium phosphate buffer pH 7.5. The peptide  
10 C-terminal thioester reacted to form a single  
11 product that eluted earlier than the starting  
12 thioester peptide as analysed by reverse phase HPLC.  
13 This material corresponded to the expected N-methoxy  
14 peptide amide as determined by ESMS: observed mass =  
15 3070 Da, expected mass 3068 Da. The kinetics of the  
16 reaction were monitored using reverse phase HPLC  
17 (Table II). The peptide C-terminal thioester was  
18 converted to the corresponding N-methoxy peptide  
19 amide derivative in a clean fashion with no side-  
20 product formation, with the reaction 75% complete  
21 after 24 h. Under these conditions no thioester  
22 hydrolysis was observed.

23

Reactant	Concentration	pH	Percentage product formation with time				
			1hr	2hr	5hr	24hr	72hr

NH <sub>2</sub> OCH <sub>3</sub>	100 mM	7.5	-	7.5	28	76	
----------------------------------	--------	-----	---	-----	----	----	--

1 Table II

2

3 When the reaction was repeated under the same  
4 conditions but with 10 mM *O*-methylhydroxylamine  
5 replacing 100 mM *O*-methylhydroxylamine, the reaction  
6 rate was slower. However, after 72h, 88% of the  
7 starting C-terminal thioester peptide had reacted.  
8 Under these conditions side-product formation was  
9 observed, in addition to the desired reaction  
10 product formation. Even so, after 72h, 30-40% of the  
11 reaction product was estimated to be the desired  
12 ligation reaction product (N-methoxy peptide amide)  
13 from HPLC analysis of the reaction mixture.

14

15 The reaction of *O*-methylhydroxylamine with  
16 recombinant C-terminal thioester proteins was also  
17 investigated. Recombinant Grb2-SH2 was generated as  
18 the C-terminal mercaptoethanesulfonic acid  
19 thioester derivative, through thiol mediated  
20 cleavage of the fusion protein Grb2-SH2 - GyrA  
21 intein - CBD, as described in Example 2. This  
22 recombinant C-terminal thioester protein was reacted  
23 with 100mM *O*-methylhydroxylamine at pH 7.5. Analysis  
24 of the reaction mixture after 18h by HPLC and ESMS  
25 showed that all of the C-terminal thioester protein

1 had been completely converted into two protein  
2 species. These two protein derivatives corresponded  
3 to the desired ligation reaction product, namely  
4 Grb2-SH2 C-terminal N-methoxy amide (expected mass  
5 12067 Da; observed mass 12067 Da), and an oxidised  
6 form of the desired reaction product (observed mass  
7 12084 Da). No side products corresponding to  
8 hydrolysis of the C-terminal protein thioester were  
9 observed. Thus all of the C-terminal thioester  
10 recombinant protein had chemoselectively ligated  
11 with *O*-hydroxylamine, via an amide bond forming  
12 reaction specifically at the C-terminus of the  
13 protein. i.e. the reaction afforded site-specific C-  
14 terminal labelling of the recombinant protein.

15

16

17

18 **Example 2- Generation of recombinant C-terminal**  
19 **hydrazide Grb2 SH2 protein.**

20

21 To investigate (i) the ability to generate  
22 recombinant C-terminal hydrazide proteins through  
23 the selective cleavage of protein - intein fusions  
24 with hydrazine, and (ii) their subsequent use in  
25 ligation / labelling reactions, the SH2 domain of  
26 the adapter protein Grb2 was chosen as a model  
27 system.

28

29 Sequence of human Grb2 SH2 domain



1 HPW FFGKIPRAKA EEMLSKQRHD GAFLIRESES APGDFSLSVK  
2 FGNDVQHFKV LRDGAGKYFL WVKFNSLNE LVDYHRSTSV  
3 SRNQQIFLRD IEQVPQQPT  
4

5 *Expression of Grb2-SH2 domain - GyrA intein fusion.*  
6 The DNA sequence encoding the SH2 domain of human  
7 Grb2 appended at its C-terminus with an extra  
8 glycine residue was cloned into the pTXB1 expression  
9 plasmid (NEB). This vector pTXB1<sub>Grb2-SH2 (Gly)</sub> encodes  
10 for a fusion protein whereby the SH2 domain of Grb2  
11 is linked via a glycine residue to the N-terminus of  
12 the GyrA intein, which is in turn fused to the N-  
13 terminus of a chitin binding domain region (CBD).  
14 *E. coli* cells were transformed with this plasmid and  
15 grown in LB medium to mid log phase and protein  
16 expression induced for 4h at 37°C with 0.5 mM IPTG.  
17 After centrifugation the cells were re-suspended in  
18 lysis buffer (0.1 mM EDTA, 250 mM NaCl, 5% glycerol,  
19 1 mM PMSF, 25 mM HEPES, pH 7.4) and lysed by  
20 sonication. The soluble fraction was loaded onto a  
21 chitin column pre-equilibrated in lysis buffer. The  
22 column was then washed with wash buffer (1 mM EDTA,  
23 250 mM NaCl, 0.1% Triton-X 100, , 25 mM HEPES, pH  
24 7.0) to yield purified Grb2-SH2 - GyrA-CBD  
25 immobilised on chitin beads (Figure 7).  
26

27 *Generation of Grb2-SH2 C-terminal thioesters by*  
28 *thiol induced cleavage of the Grb2-SH2 - GyrA intein*  
29 *fusion.*

30 To ascertain that the intein domain within the  
31 protein was functional the fusion protein was  
32 exposed to thiols to assess the extent of cleavage

1 via transthioesterification. Chitin beads containing  
2 immobilised Grb2-SH2 - GyrA-CBD were equilibrated  
3 into 200 mM NaCl, 200 mM phosphate buffer pH 7.4.  
4 Dithiothreitol (DTT) or 2-mercaptoethanesulfonic  
5 acid (MESNA) were then added to the beads in 200 mM  
6 NaCl, 200 mM phosphate buffer pH 7.4 to give a 50%  
7 slurry with a final thiol concentration of 100 mM or  
8 120 mM respectively. The mixtures were then rocked  
9 at room temperature and aliquots analysed by SDS-  
10 PAGE. After 48 hours the supernatants from the  
11 reactions were isolated and subsequently analysed by  
12 HPLC and ESMS.

13  
14 Treatment of Grb2-SH2 - GyrA intein - CBD fusion  
15 with both DTT and MESNA resulted in cleavage of the  
16 fusion protein into two protein species (Figure 7).  
17 The molecular size of the two fragments corresponds  
18 to that of the Grb2 - SH2 and the GyrA - intein  
19 fusion, indicative that cleavage has taken place at  
20 the SH2 - intein junction. Cleavage of the precursor  
21 fusion protein liberated the SH2 domain into the  
22 supernatant while the GyrA intein-CBD portion  
23 remained immobilized on the chitin beads. After  
24 cleavage with both DTT or MESNA, ESMS analysis of  
25 the supernatants confirmed that the Grb2-SH2 was  
26 generated as either the expected DTT or MESNA C-  
27 terminal thioester derivatives respectively.

28  
29 Expected mass of Grb2-SH2 DTT - C-terminal thioester  
30 = 12173.9 Da; observed mass 12173.5 Da. Expected  
31 mass of Grb2-SH2 MESNA - C-terminal thioester =  
32 12162.0 Da; observed mass 12163.0 Da.

1  
2           *Generation of Grb2-SH2 C-terminal hydrazide by*  
3 *hydrazine induced cleavage of the Grb2-SH2 - GyrA*  
4 *intein fusion.*

5  
6   The inventors hypothesised that the thioester  
7   linkage between Grb2-SH2 and the GyrA intein in the  
8   precursor fusion protein is cleaved with hydrazine.  
9   The chemoselective reaction of hydrazine, at the  
10   thioester moiety linking Grb2 SH2 to the intein,  
11   would liberate the Grb2-SH2 domain into the  
12   supernatant as its corresponding C-terminal  
13   hydrazide derivative. Chitin beads containing  
14   immobilised Grb2-SH2 - GyrA-CBD were therefore  
15   equilibrated into 200 mM NaCl, 200 mM phosphate  
16   buffer pH 7.4 and hydrazine monohydrate added in the  
17   same buffer to give a 50% slurry with a final  
18   hydrazine concentration of 200 mM. The mixture was  
19   then rocked at room temperature and analysed by SDS-  
20   PAGE (Figure 8). After 20 hours the supernatant was  
21   removed and analysed by HPLC and ESMS.

22  
23   Treatment of Grb2-SH2 - GyrA intein - CBD fusion  
24   with hydrazine resulted in cleavage of the fusion  
25   protein into two species. The molecular size of the  
26   two fragments as analysed by SDS-PAGE corresponded  
27   to Grb2 - SH2 and the GyrA - intein fusion,  
28   indicative that cleavage has taken place at the  
29   unique thioester linkage between the SH2 and intein  
30   domains. Cleavage of the precursor fusion protein  
31   liberated the SH2 domain into the supernatant while  
32   the GyrA intein-CBD portion remained immobilized on

1 the chitin beads. HPLC and ESMS analysis of the  
2 cleavage supernatant confirmed that a single protein  
3 species was generated that corresponds to the C-  
4 terminal hydrazide derivative of Grb2-SH2. Expected  
5 mass of Grb2-SH2 C-terminal hydrazide = 12051.7 Da;  
6 observed mass 12053.0 Da. (Figure 9)

7  
8 After 20 h of reaction Grb2-SH2 C-terminal hydrazide  
9 was isolated from the supernatant by either (i)  
10 using RPHPLC followed by lyophilisation or (ii) by  
11 gel filtration. In this later approach the Grb2-SH2  
12 C-terminal hydrazide reaction solution was loaded  
13 onto a superdex peptide column (Amersham  
14 Biosciences) and eluted with a running buffer of 50  
15 mM sodium acetate pH 4.5. This yielded a solution of  
16 purified Grb2-SH2 C-terminal hydrazide in 50 mM  
17 sodium acetate pH 4.5. This solution was  
18 concentrated using a centricon filter (3000 MWCO),  
19 then snap frozen and stored at -20°C until use.

20  
21 A sample of the purified and lyophilised Grb2-SH2 C-  
22 terminal hydrazide (100 µg) was treated with the  
23 protease Lys-C (5 µg) in 100mM ammonium bicarbonate  
24 buffer pH 8.2 (100 µL). After incubating at 30°C  
25 overnight the reaction was lyophilised and analysed  
26 by MALDI mass spectrometry. The observed mass of the  
27 C-terminal proteolytic fragment

28 (FNSLNELVDYHRSTSVSRNQQIFLRDIEQVPQQPTG) corresponds  
29 to that of the desired C-terminal hydrazide

1 derivative (expected mass of C-terminal hydrazide  
2 proteolytic fragment 4229 Da; observed mass 4231  
3 Da)  
4  
5

6 **Example 3- Generation of recombinant C-terminal**  
7 **hydrazide maltose binding protein.**  
8

9 As a further demonstration of the described  
10 approach, for generating recombinant C-terminal  
11 hydrazide proteins through the selective cleavage of  
12 protein - intein fusions with hydrazine, the  
13 generation of the C-terminal hydrazide derivative of  
14 maltose binding protein (MBP) was investigated.  
15

16 Sequence of human MBP used

17 M K I E E G K L V I W I N G D K G Y N G L A E V G K  
18 K F E K D T G I K V T V E H P D K L E E K F P Q V A  
19 A T G D G P D I I F W A H D R F G G Y A Q S G L L A  
20 E I T P D K A F Q D K L Y P F T W D A V R Y N G K L  
21 I A Y P I A V E A L S L I Y N K D L L P N P P K T W  
22 E E I P A L D K E L K A K G K S A L M F N L Q E P Y  
23 F T W P L I A A D G G Y A F K Y E N G K Y D I K D V  
24 G V D N A G A K A G L T F L V D L I K N K H M N A D  
25 T D Y S I A E A A F N K G E T A M T I N G P W A W S  
26 N I D T S K V N Y G V T V L P T F K G Q P S K P F V  
27 G V L S A G I N A A S P N K E L A K E F L E N Y L L  
28 T D E G L E A V N K D K P L G A V A L K S Y E E E L  
29 A K D P R I A A T M E N A Q K G E I M P N I P Q M S  
30 A F W Y A V R T A V I N A A S G R Q T V D E A L K D

1     A Q T N S S S N N N N N N N N N L G I E G R G T L  
2     E G

3

4

5     *Expression of MBP - Sce VMA intein fusion.*

6     The expression vector pMYB5 (New England Biolabs)  
7     encodes for a fusion protein comprising maltose  
8     binding protein (sequence above) fused N-terminal to  
9     the Sce VMA intein, which is in turn fused to the N-  
10    terminus of a chitin binding domain (CBD) to  
11    facilitate purification.

12

13    *E. coli* cells were transformed with this plasmid and  
14    grown in LB medium to mid log phase and protein  
15    expression induced for 4h at 37°C with 0.5 mM IPTG.  
16    After centrifugation the cells were re-suspended in  
17    lysis buffer (0.1 mM EDTA, 250 mM NaCl, 5% glycerol,  
18    1 mM PMSF, 25 mM HEPES, pH 7.4) and lysed by  
19    sonication. The soluble fraction was loaded onto a  
20    chitin column pre-equilibrated in lysis buffer. The  
21    column was then washed with wash buffer (1 mM EDTA,  
22    250 mM NaCl, 0.1% Triton-X 100, , 25 mM HEPES, pH  
23    7.0) to yield the purified fusion protein (MBP-VMA-  
24    CBD) immobilised on chitin beads.

25

26    *Generation of MBP C-terminal thioesters by thiol*  
27    *induced cleavage of the MBP - VMA- intein fusion*  
28    *protein.*

29

30    To ascertain that the intein domain within MBP-VMA-  
31    CBD was functional, the fusion protein was exposed  
32    to 2-mercaptoethanesulfonic acid (MESNA) to assess

1 the extent of cleavage via transthioesterification.  
2 Chitin beads containing immobilised MBP-VMA-CBD were  
3 equilibrated into 200 mM NaCl, 200 mM phosphate  
4 buffer pH 7.4. MESNA was then added to the beads in  
5 200 mM NaCl, 200 mM phosphate buffer pH 7.4 to give  
6 a 50% slurry with a final thiol concentration of 120  
7 mM. The mixture was then rocked at room temperature  
8 and aliquots analysed by SDS-PAGE. After 48 hours  
9 the supernatants from the reactions were isolated  
10 and subsequently analysed by HPLC and ESMS.

11  
12 Treatment of MBP-VMA-CBD fusion with MESNA results  
13 in cleavage of the fusion protein into two protein  
14 species. The molecular size of the two fragments  
15 corresponds to that of the MBP and the VMA-CBD  
16 portion, indicative that cleavage has taken place at  
17 the MBP - VMA intein junction. Cleavage of the  
18 precursor fusion protein liberates MBP into the  
19 supernatant while the VMA-CBD portion remains  
20 immobilized on the chitin beads. This was confirmed  
21 by ESMS analysis of the cleavage supernatant, which  
22 contained one protein species. Expected mass of MBP  
23 C-terminal MESNA thioester 43064 Da; observed mass  
24 43098 Da.

25  
26 *Generation of MBP C-terminal hydrazide by hydrazine*  
27 *induced cleavage of the MBP-VMA intein fusion*  
28 *protein.*

29  
30 Chitin beads containing immobilised MBP-VMA-CBD were  
31 equilibrated into 200 mM NaCl, 200 mM phosphate  
32 buffer pH 7.4 and hydrazine monohydrate added in the

1 same buffer to give a 50% slurry with a final  
2 hydrazine concentration of 200 mM. The mixture was  
3 then rocked at room temperature and analysed by SDS-  
4 PAGE and by HPLC and ESMS.

5  
6 After 20 h of reaction MBP C-terminal hydrazide was  
7 isolated from the supernatant by either (i) using  
8 RPHPLC followed by lyophilisation or (ii) by gel  
9 filtration. In this later approach the MBP C-  
10 terminal hydrazide reaction solution was loaded onto  
11 a superdex peptide column (Amersham Biosciences) and  
12 eluted with a running buffer of 50 mM sodium acetate  
13 buffer pH 4.5. This yielded a solution of purified  
14 MBP C-terminal hydrazide in 50 mM sodium acetate  
15 buffer pH 4.5. This protein solution was  
16 concentrated using a centricon filter (3000 MWCO),  
17 then snap frozen and stored at -20°C until use.

18  
19 Treatment of MBP-VMA-CBD fusion with hydrazine  
20 results in cleavage of the fusion protein into two  
21 species. The molecular size of the two fragments as  
22 analysed by SDS-PAGE corresponds to MBP and the VMA-  
23 CBD portion, indicative that cleavage has taken  
24 place at the unique thioester linkage between the  
25 MBP - VMA intein domain. Cleavage of the precursor  
26 fusion protein liberates MBP into the supernatant,  
27 while the VMA-CBD portion remains immobilized on the  
28 chitin beads. HPLC and ESMS analysis of the cleavage  
29 supernatant confirms that a single protein species  
30 is generated with an observed mass of 42988 Da. The  
31 expected mass difference between the C-terminal  
32 MESNA thioester derivative of a protein and its



1 corresponding C-terminal hydrazide is 111 Da. The  
2 observed mass of the C-terminal MESNA thioester of  
3 MBP was found to be 43098 Da. Thus the product from  
4 the hydrazine cleavage of MBP-VMA- CBD is 110 Da  
5 lower, indicating that the desired C-terminal  
6 hydrazide derivative of MBP had been formed.

7

8 **Example 4- Ligation of aldehyde and ketone**  
9 **containing peptides and labels to recombinant C-**  
10 **terminal hydrazide containing proteins: Ligation of**  
11 **a synthetic peptide c-myc to recombinant Grb2 SH2**  
12 **domain.**

13

14 The inventors hypothesised that recombinant protein  
15 C-terminal hydrazides, generated by hydrazine  
16 treatment of the corresponding intein fusion  
17 precursor, can be site-specifically modified by  
18 chemoselective ligation with aldehyde and ketone  
19 containing peptides and labels. To demonstrate such  
20 an approach, the ability of a synthetic ketone  
21 containing peptide to ligate with the Grb2-SH2 C-  
22 terminal hydrazide generated above was investigated.  
23 A synthetic peptide corresponding to the c-myc  
24 epitope sequence was synthesised GEQKLISEEDL-NH<sub>2</sub>,  
25 whereby pyruvic acid was coupled to the amino  
26 terminus of the peptide as the last step of the  
27 assembly. This peptide (designated CH<sub>3</sub>COCO-myc) was  
28 purified to > 95% purity by RPHPLC and lyophilised  
29 (ESMS expected monoisotopic mass 1328.6 Da; observed  
30 mass 1328.6 Da).

31

1 A sample of CH<sub>3</sub>COCO-myc peptide was dissolved in 100  
2 mM sodium acetate buffer pH 4.5 to give a 4 mM  
3 peptide concentration. This peptide solution (100  
4 µL) was then added to an aliquot of lyophilised  
5 Grb2-SH2 C-terminal hydrazide protein (~ 250 µg) and  
6 the reaction monitored by SDS-PAGE (Figure 10) As a  
7 control CH<sub>3</sub>COCO-myc was also incubated with  
8 Cytochrome C, a protein of similar same size to  
9 Grb2-SH2 but absent of a hydrazide functionality.  
10  
11 SDS-PAGE analysis shows that CH<sub>3</sub>COCO-myc peptide  
12 has indeed ligated with Grb2-SH2 C-terminal  
13 hydrazide, as indicated by the conversion of Grb2-  
14 SH2 C-terminal hydrazide into a protein species of  
15 a higher molecular weight (approximately 1000-2000  
16 Da higher). The reaction is virtually complete after  
17 24 h and the reaction product appears to be stable.  
18 On the other hand, there was no observable change to  
19 Cytochrome C with time i.e no ligation, establishing  
20 that the ligation reaction is occurring at the C-  
21 terminal hydrazide functionality of Grb2-SH2.  
22  
23 After 96 h of reaction the product from the Grb2-SH2  
24 ligation reaction was isolated by HPLC and  
25 characterised by ESMS. Chemoselective ligation of  
26 CH<sub>3</sub>COCO-myc to Grb2-SH2 C-terminal hydrazide via  
27 hydrazone bond formation would give a product of  
28 expected mass 13363.7 Da. The observed product mass  
29 was 13364.1 Da indicting that the desired ligation  
30 product had been formed.  
31

**Example 5- Ligation of aldehyde and ketone  
containing peptides and labels to recombinant C-  
terminal hydrazide containing proteins: Fluorescein  
labelling of Grb2-SH2.**

In this example the recombinant C-terminal hydrazide derivative of Grb2-SH2, generated through hydrazine cleavage of the precursor intein fusion protein, was reacted with a ketone containing derivative of fluorescein to afford site-specific fluorescent labelling of the protein.

To facilitate fluorescent labelling of C-terminal hydrazide recombinant proteins using the described approach, the fluorophore needs to contain the appropriate reactive group for ligation, namely an aldehyde or ketone functionality. To this end a derivative of fluorescein was synthesized containing a pyruvoyl moiety. Initially, Fmoc-Lys(Mtt)-OH was coupled to a rink amide resin, and the Mtt group removed using standard procedures (1% TFA, 4% triisopropylsilane in dichloromethane). 5(6)-carboxyfluorescein was then coupled to the lysine  $\epsilon$ -amino group. The Fmoc group was then removed and pyruvic acid coupled to the free  $\alpha$ -amino group of the lysine. After cleavage from the resin, the desired fluorescein derivative [designated CH<sub>3</sub>COCO-Lys(Fl), see Figure 11] was purified to > 95% purity by RPHPLC and lyophilised (ESMS, expected monoisotopic mass 576.2 Da; observed monoisotopic mass 576.0 Da).

1 To establish the reactivity of CH<sub>3</sub>COCO-Lys(F1) with  
2 C-terminal hydrazide peptides and proteins, the  
3 reaction of CH<sub>3</sub>COCO-Lys(F1) with a small synthetic  
4 C-terminal hydrazide peptide SLAYG-NHNH<sub>2</sub> was  
5 investigated. A sample of CH<sub>3</sub>COCO-Lys(F1) and SLAYG-  
6 NHNH<sub>2</sub> peptide were co-dissolved in 100 mM sodium  
7 acetate buffer pH 4.5 to give final concentrations  
8 of 0.3 mM and 2 mM respectively. After 20 h  
9 incubation at room temperature, the reaction was  
10 deemed complete as determined by RPHPLC analysis.  
11 All the starting CH<sub>3</sub>COCO-Lys(F1) had reacted to give  
12 predominantly a single product. The mass of which  
13 corresponds to the desired ligation product, namely  
14 conjugation of the two reactants via hydrazone bond  
15 formation (ESMS expected monoisotopic mass 1079 Da;  
16 observed mass 1080 Da).

17

18 Having established the specific reaction of CH<sub>3</sub>COCO-  
19 Lys(F1) with hydrazide containing peptides, this  
20 fluorescein derivative was used for the site-  
21 specific labeling of recombinant Grb2 SH2 C-terminal  
22 hydrazide (generated through hydrazine cleavage of  
23 Grb2 SH2 - GyrA - CBD).

24

25 Two complementary methods were employed for the  
26 purification of Grb2 SH2 C-terminal hydrazide from  
27 the fusion protein cleavage reaction (Example 2).  
28 The purified protein was isolated as either a  
29 lyophilized solid or in a solution of 50 mM sodium  
30 acetate buffer pH 4.5. This latter buffer system was  
31 chosen as the pH is suited to hydrazone bond forming  
32 reactions. An aliquot of Grb2 SH2 C-terminal

1 hydrazide in 50mM sodium acetate pH 4.5 (250 µg, 200  
2 µL) was added directly to a sample of CH<sub>3</sub>COCO-  
3 Lys(Fl) to give a final concentration of fluorophore  
4 of circa 0.3 mM. The reaction was incubated at room  
5 temperature and monitored by SDS-PAGE. As a control  
6 CH<sub>3</sub>COCO-Lys(Fl) was also incubated under the same  
7 conditions with Cytochrome C, a protein of similar  
8 same size to Grb2-SH2 but absent of a hydrazide  
9 functionality.

10

11 SDS-PAGE analysis shows that CH<sub>3</sub>COCO-Lys(Fl) has  
12 indeed ligated with Grb2-SH2 C-terminal hydrazide  
13 (Figure 12) as indicated by the conversion of Grb2-  
14 SH2 C-terminal hydrazide into a single protein  
15 species with an apparent increase in molecular  
16 weight (approximately 1000-2000 Da higher). After  
17 SDS-PAGE analysis of the reactions, fluorescence  
18 imaging of the gel confirmed that the newly formed  
19 reaction product contains a fluorescein label, and  
20 that the reaction is clean, with only a single  
21 fluorescent protein product being formed (figure  
22 14). The reaction is virtually complete after 24 h  
23 and the reaction product appears to be stable under  
24 these conditions.

25

26 On the other hand there was no observable change to  
27 Cytochrome C over the time course of the experiment  
28 i.e no ligation (Figure 13) with a complete absence  
29 of the formation of any fluorescent protein products  
30 (Figure 14). Thus establishing that the ligation  
31 reaction is occurring at the C-terminal hydrazide  
32 functionality of Grb2 SH2, to yield site-specific C-

1 terminal fluorescent labelling of the recombinant  
2 protein. After 48 h of reaction, the product from  
3 the ligation reaction with Grb2 SH2 was isolated by  
4 HPLC. The mass of this product, by ESMS, confirmed  
5 the addition of one fluorescein group to the  
6 protein.

7  
8 In another example, lyophilised Grb2 SH2 C-terminal  
9 hydrazide was directly dissolved into 100 mM sodium  
10 acetate pH 4.5 and added to CH<sub>3</sub>COCO-Lys(F1). Whilst  
11 some protein precipitation was observed, the soluble  
12 fraction of the protein reacted with CH<sub>3</sub>COCO-Lys(F1)  
13 in the anticipated manner described above.

14  
15 In an alternative strategy, a lyophilized sample of  
16 Grb2 SH2 C-terminal hydrazide (250 µg) was dissolved  
17 in 40% aqueous acetonitrile containing 0.1% TFA  
18 (200 µL). This solution was then added to a sample  
19 of CH<sub>3</sub>COCO-Lys(F1) to give a final fluorophore  
20 concentration of circa 0.3 mM. The solution was  
21 incubated at room temperature and the reaction  
22 periodically analyzed. SDS-PAGE analysis showed that  
23 the labeling reaction had occurred cleanly and  
24 rapidly under these conditions (Figure 15). Grb2 SH2  
25 C-terminal hydrazide was converted into a single  
26 protein species with an apparent increased molecular  
27 weight expected for that of the desired product, and  
28 this newly formed protein was green fluorescent when  
29 visualised under a UV lamp. ESMS of the reaction  
30 product confirmed that one fluorescein molecule had  
31 been added to the protein. The reaction is virtually  
32 complete after 4 h, with prolonged incubation

1 appearing to be detrimental to the formation of the  
2 ligation product.

3

4 **Example 6- Ligation of aldehyde and ketone**  
5 **containing peptides and labels to recombinant C-**  
6 **terminal hydrazide containing proteins: Fluorescein**  
7 **labelling of MBP.**

8

9 As a further exemplification, the described approach  
10 was used for the site-specific C-terminal labeling  
11 of MBP with fluorescein. A sample (250 µg) of  
12 lyophilised recombinant MBP C-terminal hydrazide  
13 (generated through hydrazine cleavage of MBP - VMA -  
14 CBD precursor fusion protein) was dissolved in 40%  
15 aqueous acetonitrile containing 0.1% TFA (200 µL).  
16 The solution was then added to a sample of CH<sub>3</sub>COCO-  
17 Lys(Fl) to give a final fluorophore concentration of  
18 circa 0.3 mM. The reaction was then incubated at  
19 room temperature and periodically analyzed by SDS-  
20 PAGE.

21

22 SDS-PAGE analysis showed that the fluorescein  
23 labelling reaction had occurred under these  
24 conditions, as indicated by the formation of a  
25 single green fluorescent species with a molecular  
26 weight of circa 42 KDa. MALDI analysis of the  
27 reaction mixture after 48 h was consistent with the  
28 addition of one fluorescein molecule to MBP.

29

30 In summary, the present invention provides novel  
31 methods of protein ligation and protein labelling.  
32 These enable both synthetic and recombinantly

1     derived protein fragments to be efficiently joined  
2     together in a regioselective manner. This thus  
3     enables large proteins to be constructed from  
4     combinations of synthetic and recombinant fragments  
5     and allows proteins of any size to be site-  
6     specifically modified in an unprecedented manner.  
7     This is of major importance for biological and  
8     biomedical science and drug discovery when one  
9     considers that the ~ 30,000 human genes yield  
10    hundreds of thousands of different protein species  
11    through post-translational modification. Such post-  
12    translationally modified proteins cannot be accessed  
13    through current recombinant technologies.

14  
15    The application of such protein ligation techniques  
16    may be used for protein based tools, protein  
17    therapeutics and in de novo design and may open up  
18    many new avenues in biological and biomedical  
19    sciences that have hitherto not been possible.

20  
21    All documents referred to in this specification are  
22    herein incorporated by reference. Various  
23    modifications and variations to the described  
24    embodiments of the inventions will be apparent to  
25    those skilled in the art without departing from the  
26    scope and spirit of the invention. Although the  
27    invention has been described in connection with  
28    specific preferred embodiments, it should be  
29    understood that the invention as claimed should not  
30    be unduly limited to such specific embodiments.  
31    Indeed, various modifications of the described modes  
32    of carrying out the invention which are obvious to



1     those skilled in the art are intended to be covered  
2     by the present invention.

3

4